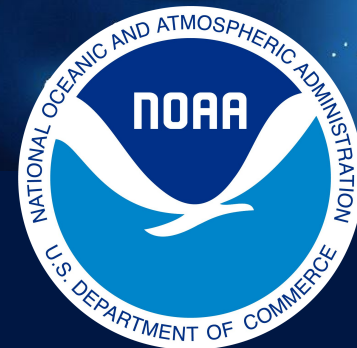40th WGNE workshop
November 3-7, *Beijing, China*

AI/ML Activities
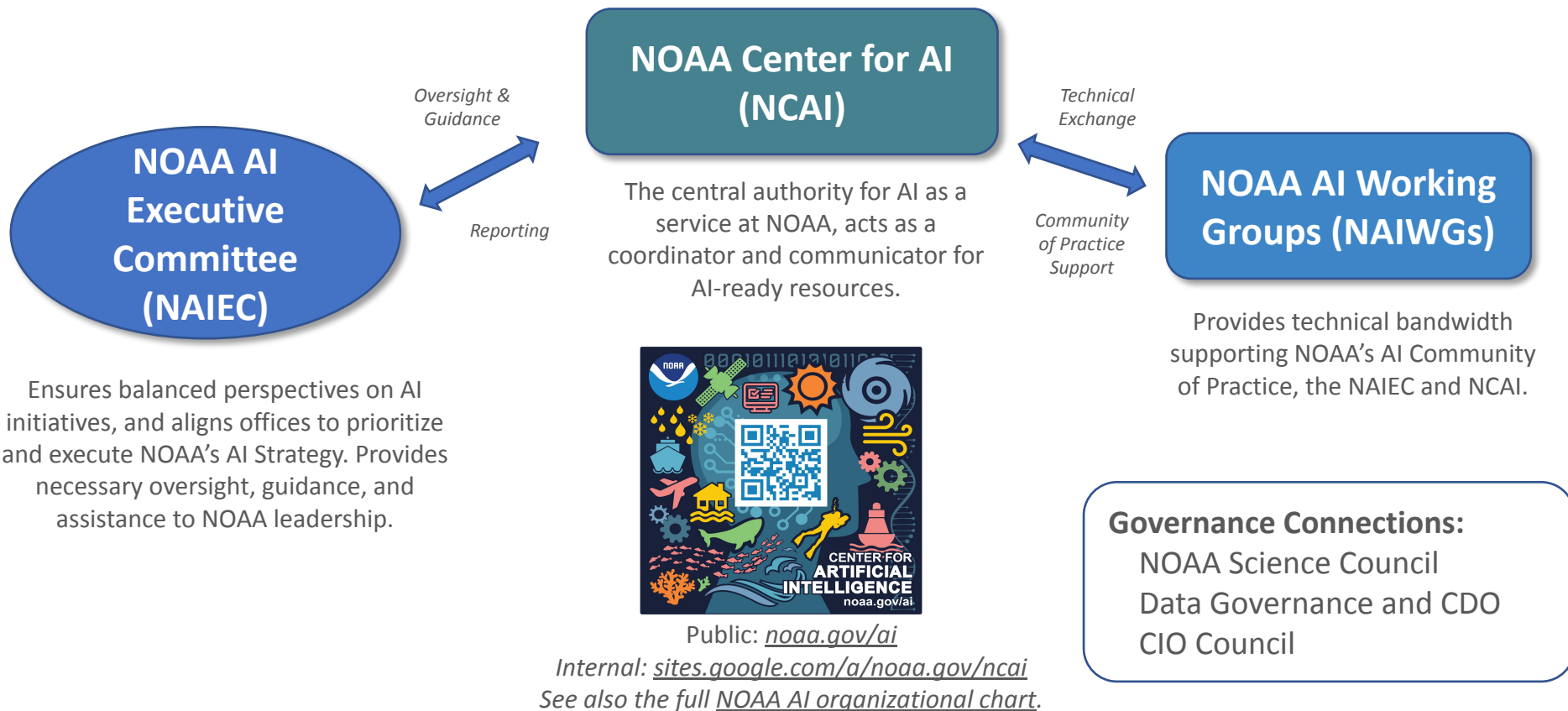**NCEP Environmental Modeling Center**

Fanglin Yang

Environmental Modeling Center
National Oceanic and Atmospheric Administration
U.S. Department of Commerce

Disclaimer: NCEP consists of nine centers, including EMC where I work. Nowadays, AI and ML activities are present in nearly every aspect of our personal and professional lives. The information I present here does not represent all NCEP or even all EMC activities; it includes only those efforts that I am personally aware of or involved in.

Acknowledgment: This presentation is made possible with contributions from many EMC developers and community collaborators. Special thanks go to Jun Wang, Linlin Cui, Azadeh Gholoubi, Zhan Zhang, Wei Li, Mary Bennett (summer Intern), and Kai Wang.

NOAA AI Governance

NOAA Center for AI (NCAI)

The central authority for AI as a service at NOAA, acts as a coordinator and communicator for AI-ready resources.

Oversight & Guidance

Reporting

Technical Exchange

Community of Practice Support

NOAA AI Executive Committee (NAIEC)

Ensures balanced perspectives on AI initiatives, and aligns offices to prioritize and execute NOAA's AI Strategy. Provides necessary oversight, guidance, and assistance to NOAA leadership.

NOAA AI Working Groups (NAIWGs)

Provides technical bandwidth supporting NOAA's AI Community of Practice, the NAIEC and NCAI.

Public: *noaa.gov/ai*
Internal: *sites.google.com/a/noaa.gov/ncai*
See also the full NOAA AI organizational chart.

Governance Connections:
 NOAA Science Council
 Data Governance and CDO
 CIO Council

Topics

- **AI Global Modeling**

- **AI regional modeling**

- **AI Observation Forecasting**

- **Miscellaneous Applications**

AI for Global Numerical Weather Prediction at NCEP

- 2024: NOAA establishes AI4NWP group. Together with NOAA Labs, EMC begins experimenting with real-time system.

- Spring 2025: NOAA Launches **EAGLE** - **E**xperimental **AI G**lobal and **L**imited-area **E**nsemble project as a means to coordinate internally and externally on advancing AI prediction.

- July 2025: NOAA Leadership encourages operational implementation, including AIGFS, AIGEFS, and HGEFS

Developing AIGFS: Training GraphCast with GDAS as input

- GraphCast: A GNN architecture with an encoder-decoder configuration (*Lam et al., 2023*) trained on 1979–2017 ERA5 data.

- Compute Node:
  - Parallel Works AWS Cloud
  - 8 H100 80GB memory GPU cores

- Training period:
  - 2021-03-21 - 2022-09-01 (4 cycles/day)

- Validation period:
  - 2022-09-01 - 2023-01-01 (4 cycles/day)

- Verification period:
  - 2023-01-01 - 2024-01-01 (2 cycles/day: 0z and 12z)

- Configuration: **Fine-tuned GraphCast against ERA5/HRES/GDAS**

- Training steps
  - Fine-tuning GC for 12, 13 and 14 Autoregressive steps using GDAS and ERA5 data

- Metrics:
  - RMSE,
  - ACC (ERA5 Climatology data were used to calculate ACC)

- Regions Evaluated:
  - Global,
  - North America,
  - Northern Hemisphere,
  - Southern Hemisphere,
  - Tropics

AIGFS Evaluation: ACC

https://www.emc.ncep.noaa.gov/users/verification/global/gfs/expr/
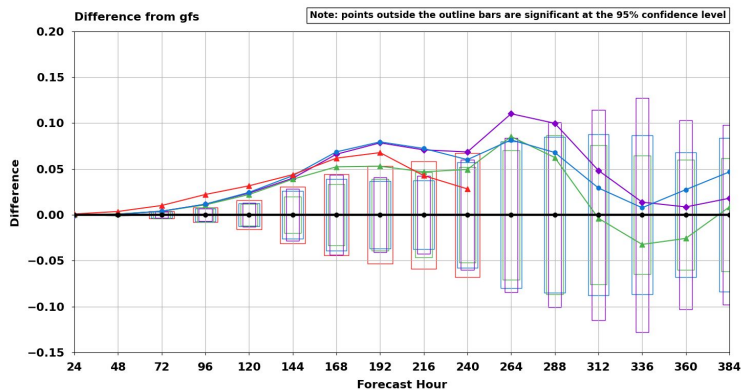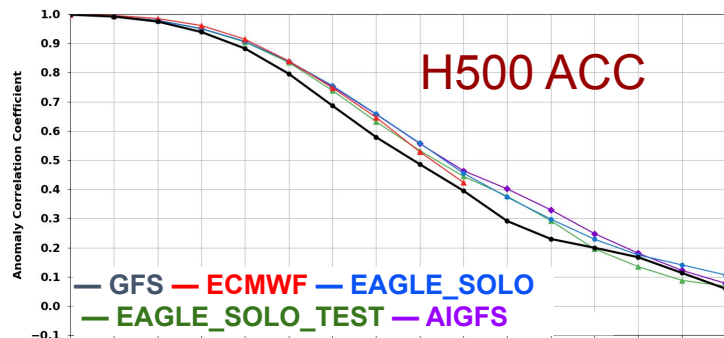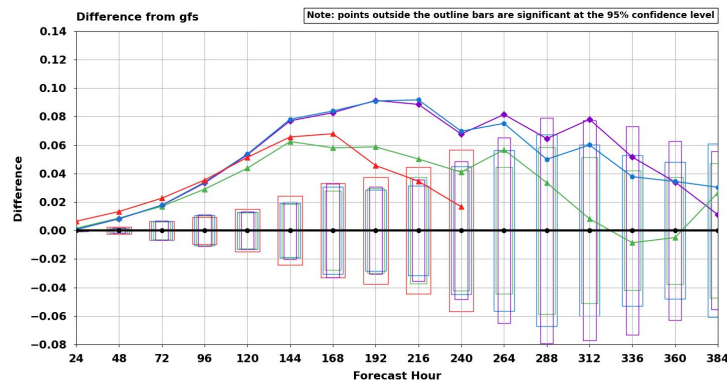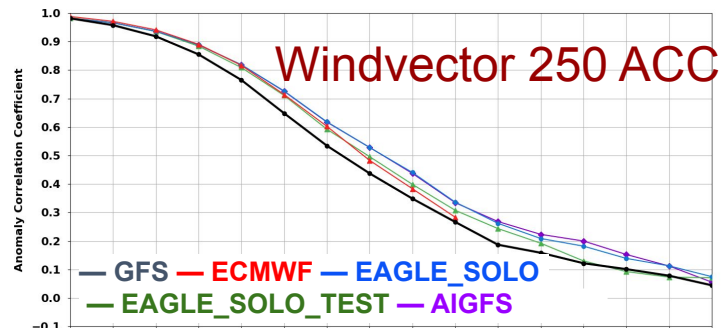
AIGFS Evaluation: Total precipitation
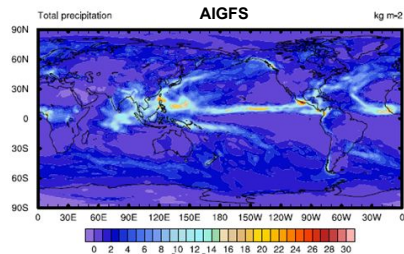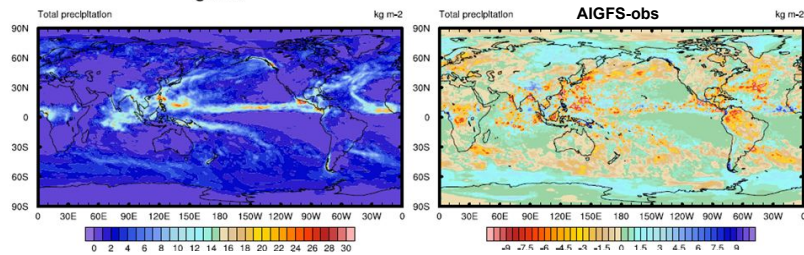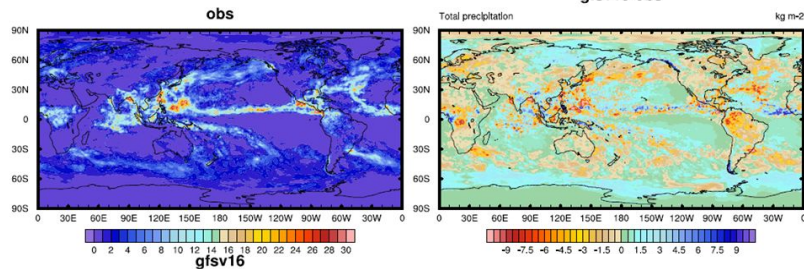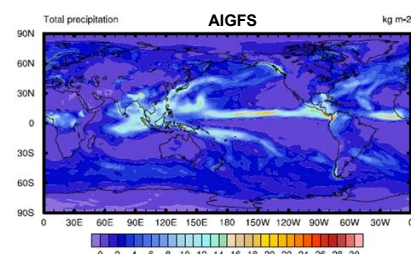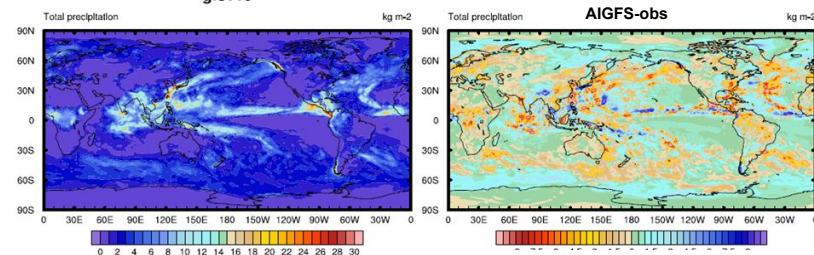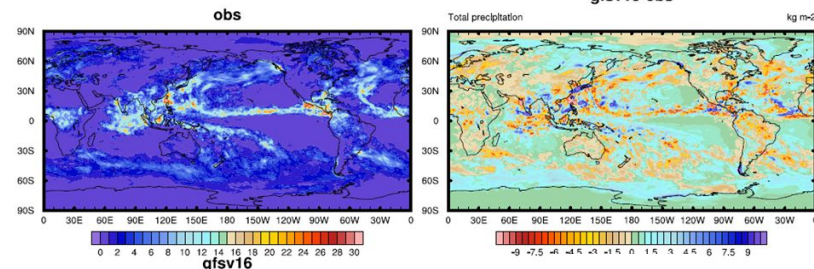
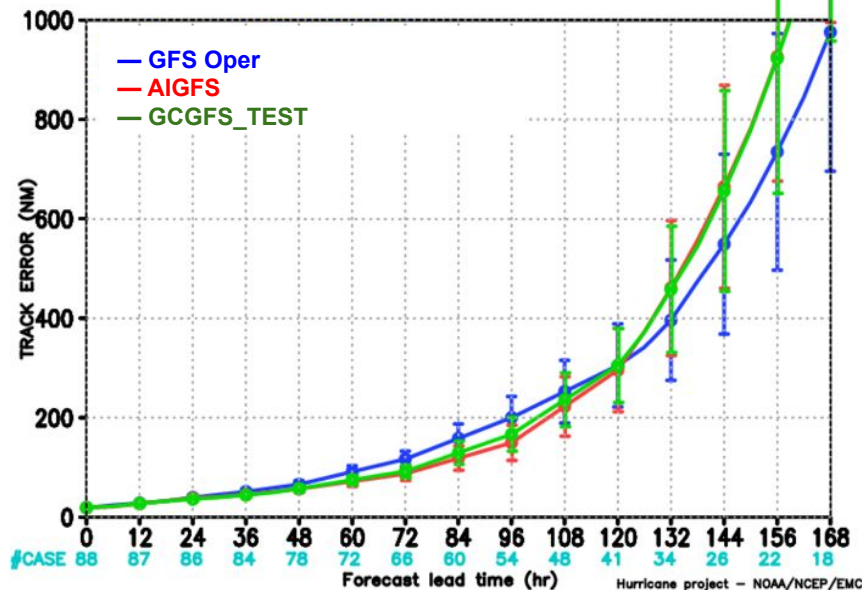- AIGFS has a reasonable total precipitation forecast, but it is much smoother.

AIGFS Evaluation: Tropical cyclones

MODEL FORECAST – TRACK ERROR (NM) STATISTICS
VERIFICATION FOR ATLANTIC BASIN 2021–2024
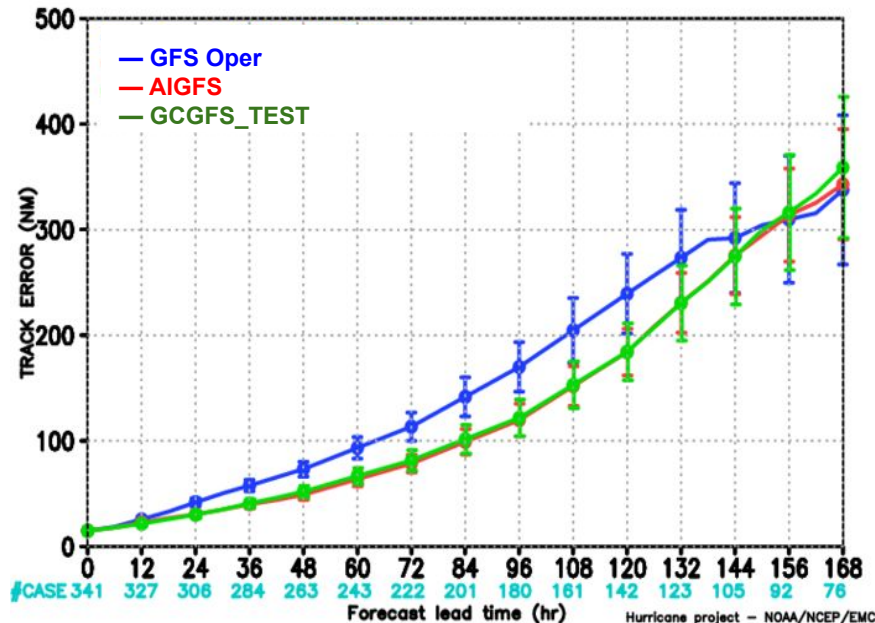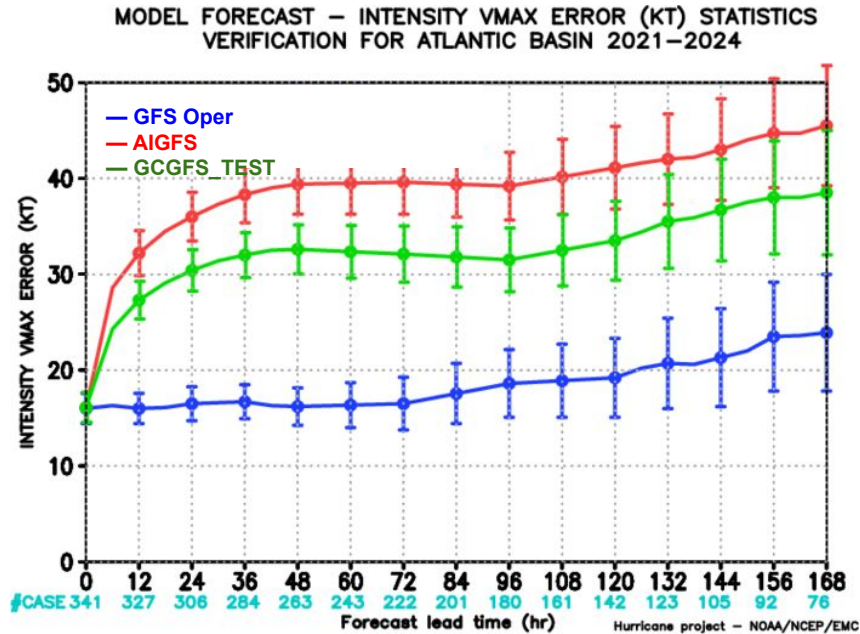
Track error in NATL

MODEL FORECAST – TRACK ERROR (NM) STATISTICS
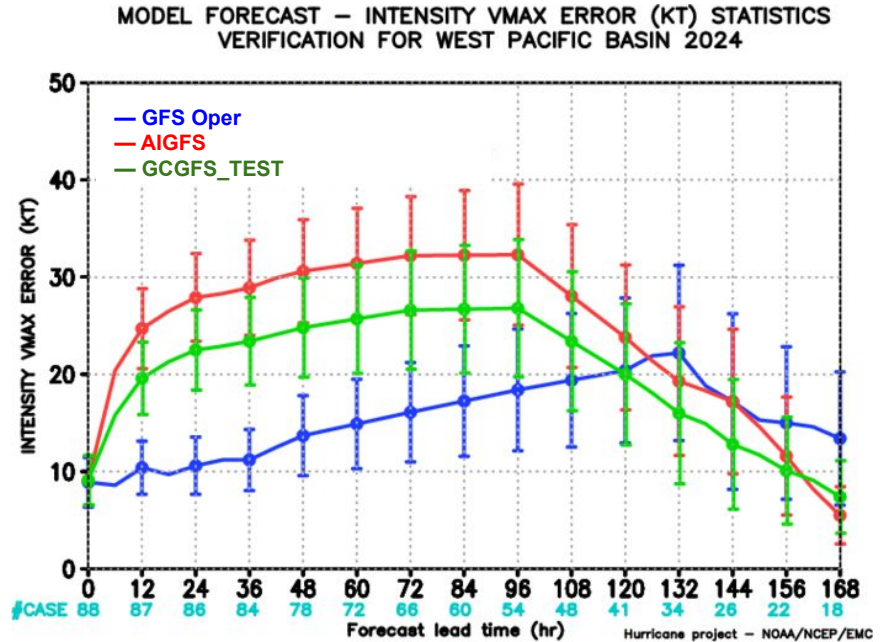VERIFICATION FOR WEST PACIFIC BASIN 2024

Track error in WPAC

- AIGFS shows better track forecasts up to day 7 in the North Atlantic basin and up to day 5 in the West Pacific basin.

# AIGFS Evaluation: Tropical cyclone intensity error

NATL basin

WPAC basin

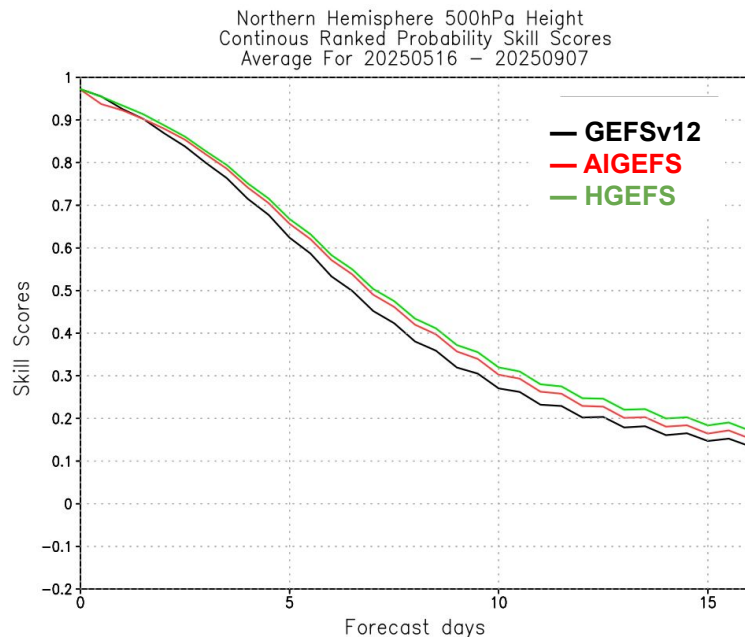- AIGFS shows worse (weak) intensity error compared to operational GFS.

Developing AIGEFS/HGEFS

- Build AIGEFS through fine-tuning and training GraphCast with GFS data as AIGEFS ensemble member weights.
- Build hybrid ensemble forecast model with AI global ensemble and the physical model GEFSv12

- Fine-tuned GraphCast:
  - GraphCast was fine-tuned at **Auto REgression** steps 12, 13, 14 with GDAS/ERA5/HRES
  - Checkpoints were picked during the training
  - A total of 18 members are picked (6 members from each AR step)

- Train GraphCast
  - Train Graphcast with GDAS data from AR steps 1 - 12
  - 12 members, each member from one AR step
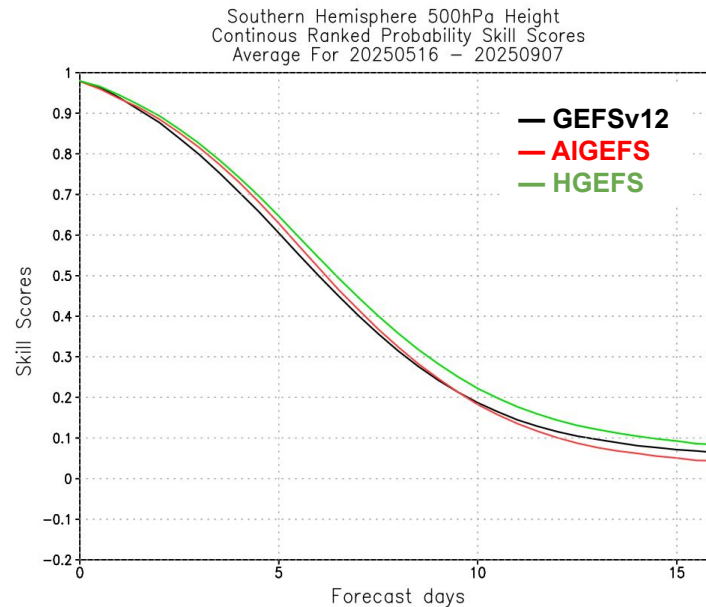- Pretrained GraphCast from Deepmind

Z500 550(green)/580(blue) dam Spaghetti maps (init time: 20251017 00Z)
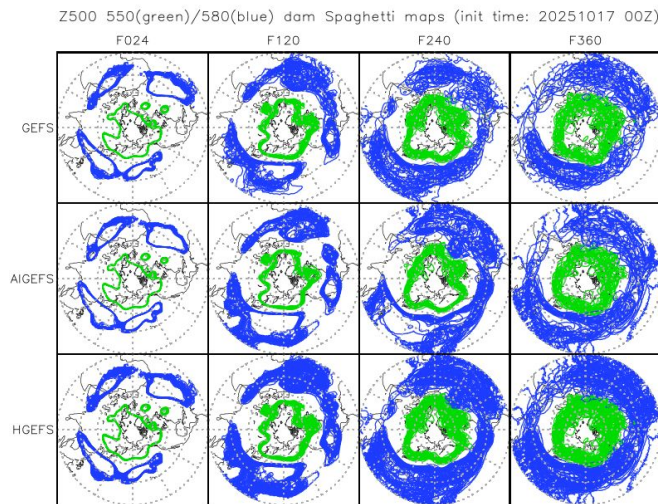
AIGEFS/HGEFS evaluation: CRPSS
111 cases, 20250516 - 20250907, reference data: GFSv16 analysis

Northern Hemisphere 500hPa Height
Continous Ranked Probability Skill Scores
Average For 20250516 - 20250907

NH: Z500 CRPSS

Southern Hemisphere 500hPa Height
Continous Ranked Probability Skill Scores
Average For 20250516 - 20250907

SH: Z500 CRPSS

- AIGEFS shows a better ensemble mean CRPSS in the northern hemisphere for lead times >1 day and in the southern hemisphere up to 9 days.
- HGEFS has a comparable ensemble mean CRPSS at short lead times and better scores at longer lead times in both hemispheres.

AIGEFS/HGEFS evaluation: RMSE, ensemble spread, errors
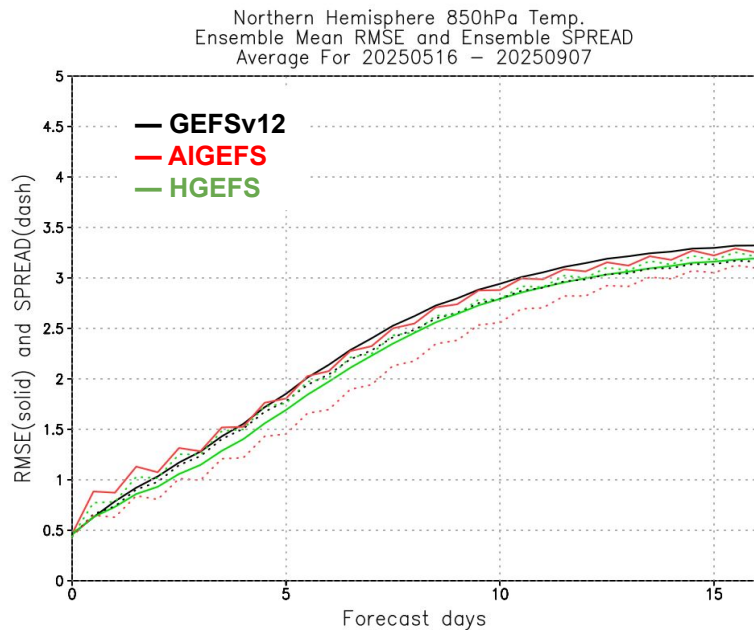

NH: T850 RMSE and ensemble spread
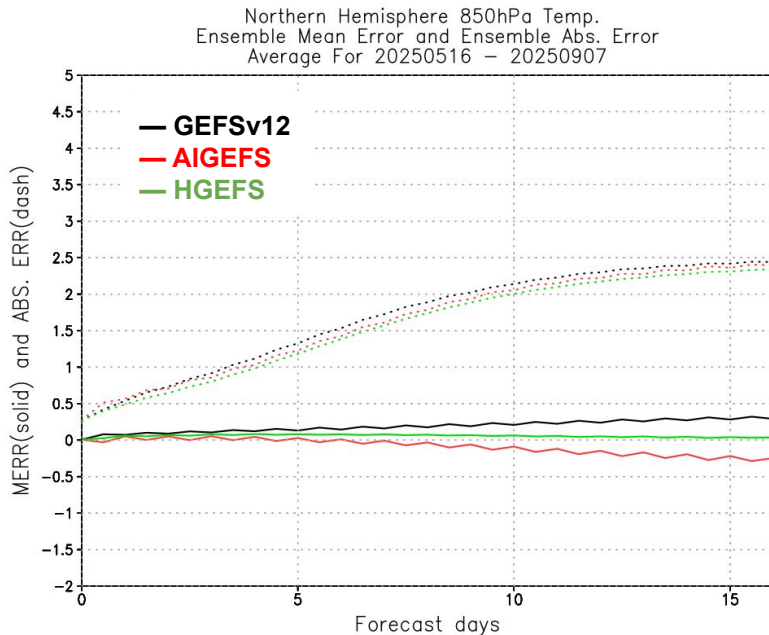
NH: T850 Ensemble mean error and abs. error

- AIGEFS has a slightly better RMSE for longer lead days, but the spread is too small. It also has a smaller abs. error and a negative mean bias.
- HGEFS is better than AIGEFS and GEFS for RMSE, ensemble spread/error ratio, abs. Error and mean error.

AIGEFS/HGEFS evaluation: tropical cyclone
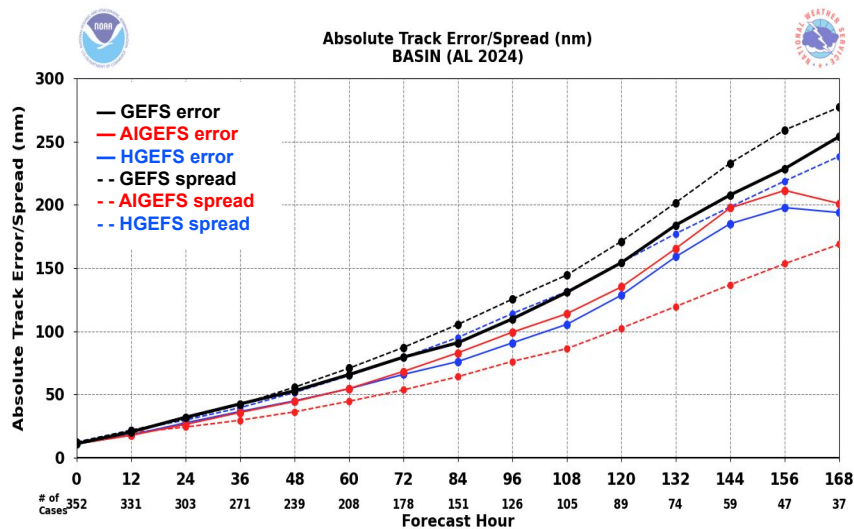
Track error: 2024 ATL Basin
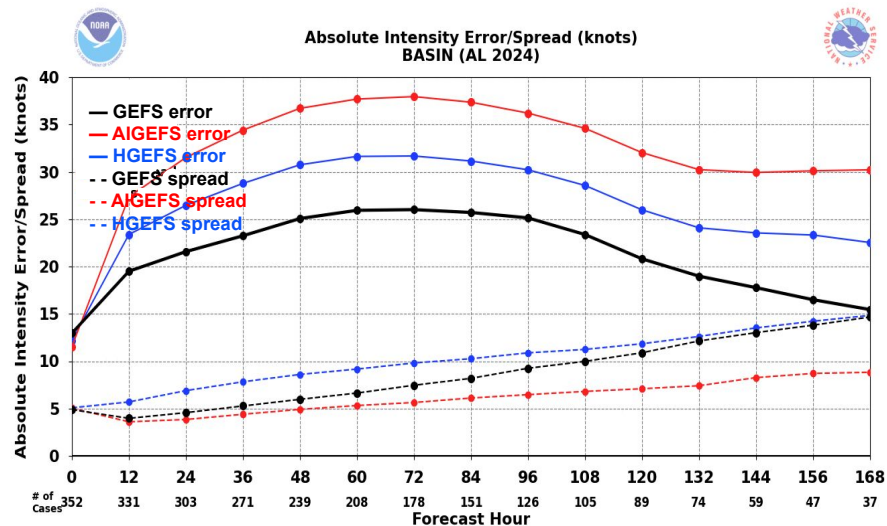
Intensity error: 2024 ATL Basin

- AIGEFS and HGEFS have a smaller tracker error than operational GEFS. The ensemble spread in AIGEFS is too small.
- AIGEFS has a largest intensity error among the three models. HGEFS has the largest ensemble spread.

AI-GFS, AI-GEFS and HGEFS Implementation Plan

30 days after US government reopens

Data Access: NOMADS, AWIPS

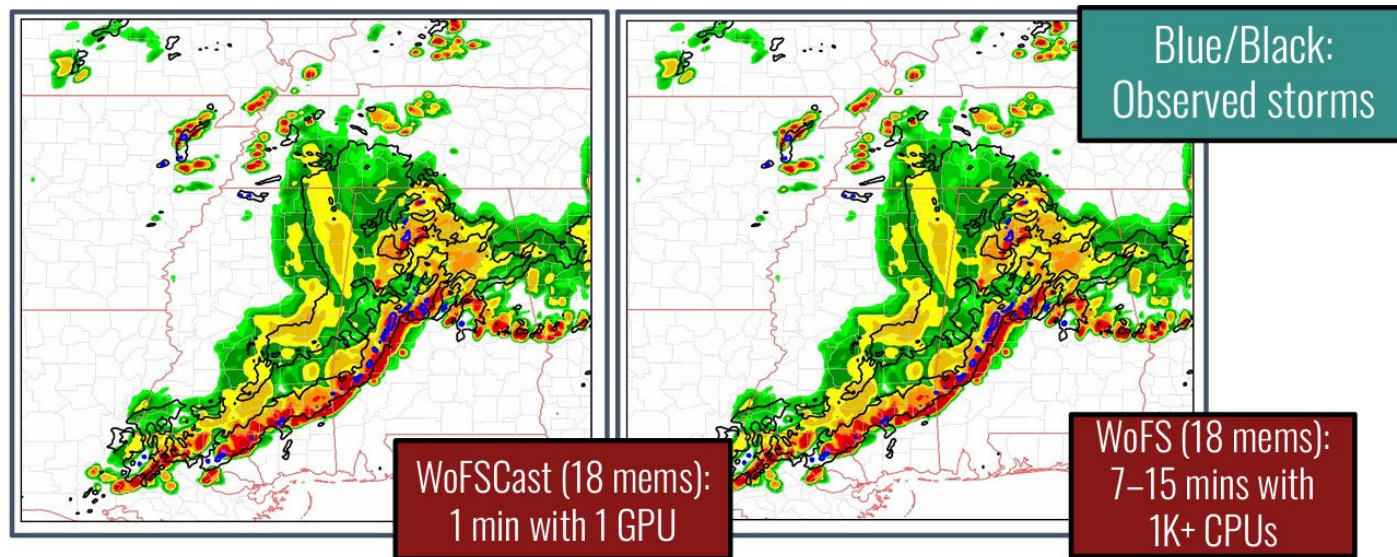**Data Viz: Coming to NWS DESI after shutdown**

AI-GFS, AI-GEFS and HGEFS Implementation Plan

- **30 days after US government reopens**
- **Data Access: NOMADS, AWIPS**
- **Data Viz: Coming to NWS DESI after shutdown**

Future Development Plans

- Use High-res GDAS analysis data for training and fine tuning
- Pressure scaling to improve prediction in the upper atmosphere
- More prognostic variables
- Add diagnostic products
- Improve hurricane intensity
- Improve ensemble spread

Regional Application: NOAA/OAR/NSSL WoFScast

WoFSCast accurately captures storm-scale evolution

Blue/Black:
Observed storms

WoFSCast (18 mems):
1 min with 1 GPU

WoFS (18 mems):
7–15 mins with
1K+ CPUs

Cory Potvin
@UIFCW

Regional Application: NOAA/OAR/GSL   HRRR-Cast

HRRR-Cast, considered a data-driven sibling to HRRR, is an **in-house AI model** trained on a three-year dataset from the physics-driven HRRR. It began as an emulator, which reproduces the output of its training model, developed with a goal of matching the skill of HRRR.

Recent refinements, including the integration of diffusion techniques enabled easy creation of ensemble forecasts through perturbations of initial conditions.
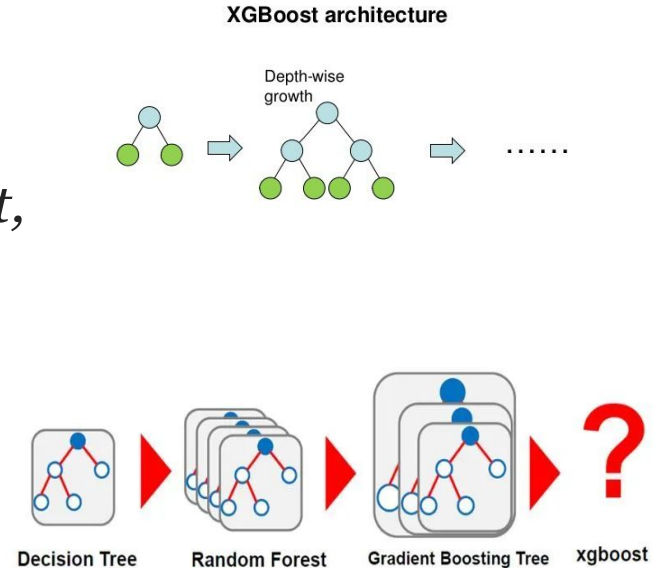

https://gsl.noaa.gov/news/hrrr-cast-unleashes-noaas-ai-weather-forecasting
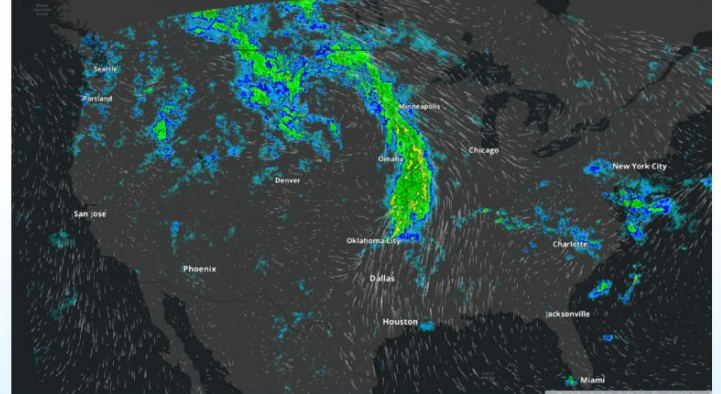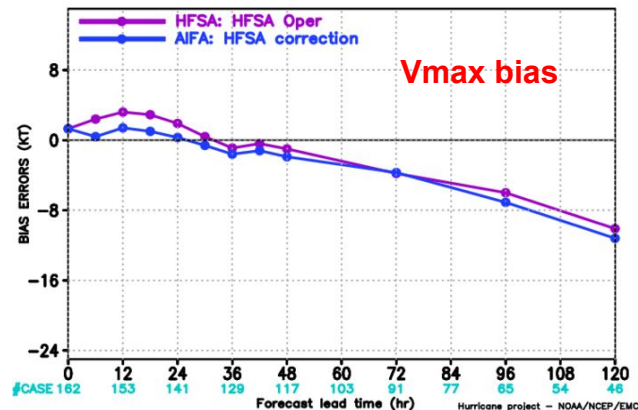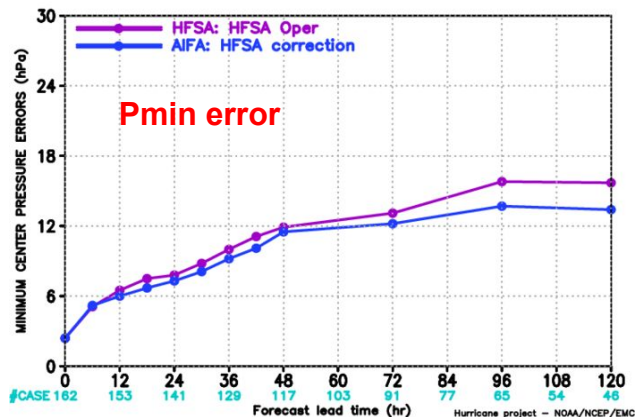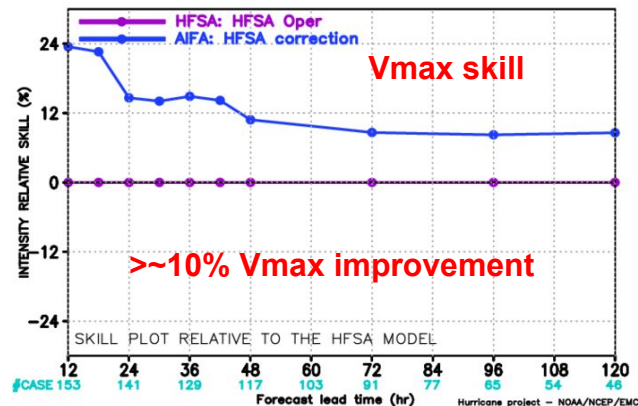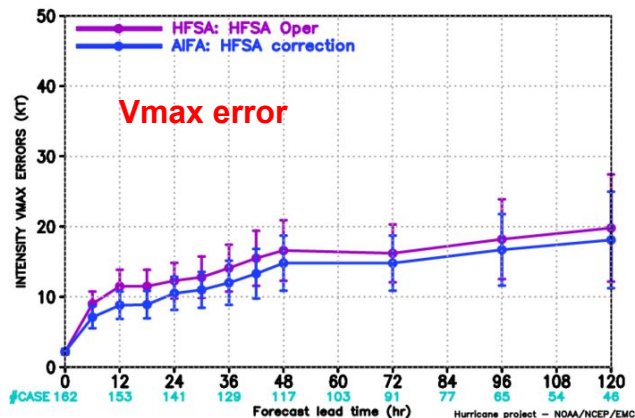
Regional Application: NOAA/NWS/EMC AI-HAFS

- ATCF based AI/ML HAFS for TC Intensity
- HAFS ensemble AI/ML emulator (in planning)
- High resolution HAFS reanalysis (in planning)

XGBoost Algorithm

- *XGBoost stands for "Extreme Gradient Boosting"*
- *XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible, and portable.*
- *It implements Machine Learning algorithms under the Gradient Boosting framework.*
- *It provides a parallel tree boosting to solve many data science problems in a fast and accurate way.*

XGBoost architecture

Depth-wise growth

......

Decision Tree   Random Forest   Gradient Boosting Tree   xgboost
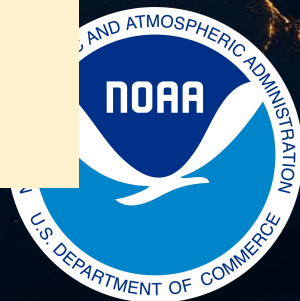
Vmax/Pmin Forecasts (AIFA vs. HFSA)

OCELOT

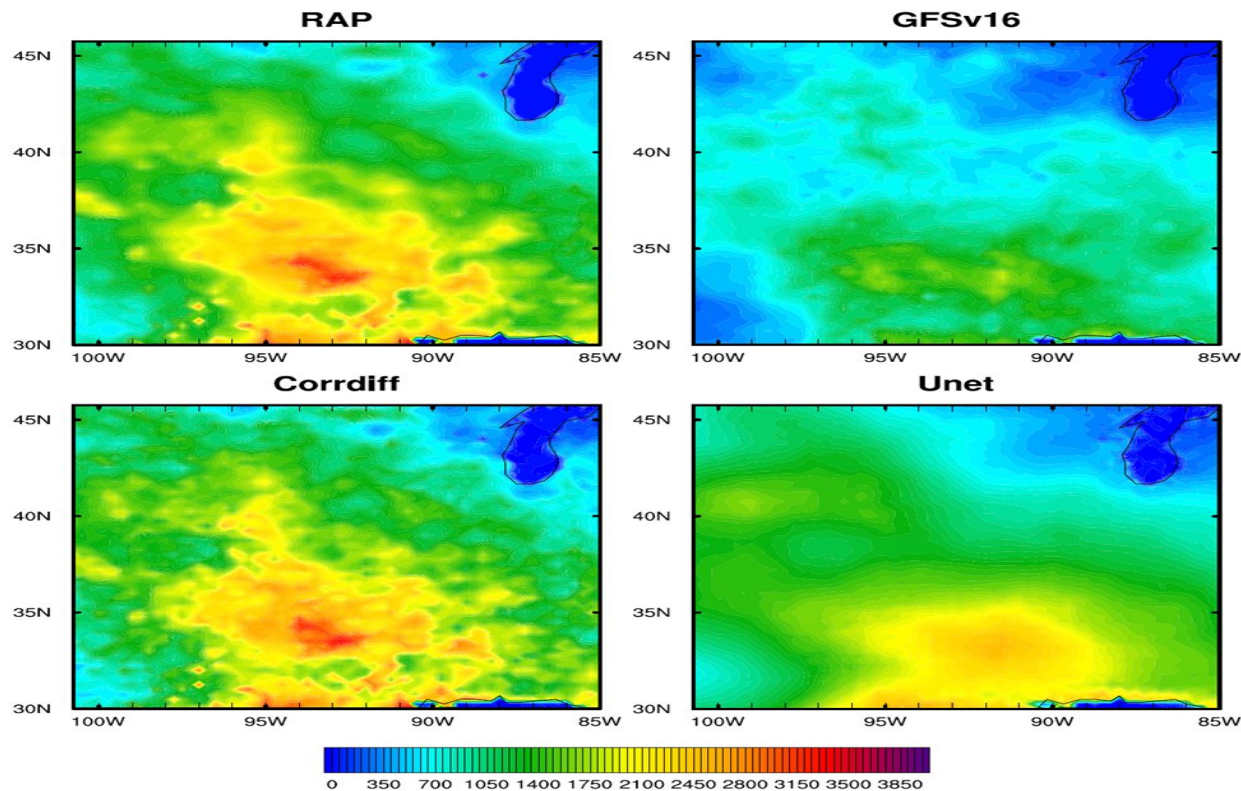## Observation Centric Estimation and Learning for Outlook Trajectories

A Graph Neural Network Framework for Multi-Source Earth Observation Forecasting

- **A GNN framework for forecasting directly from raw observations**

- **Integrates satellite + conventional data**

- **Forecasts in obs space without reliance on reanalysis or physics models**

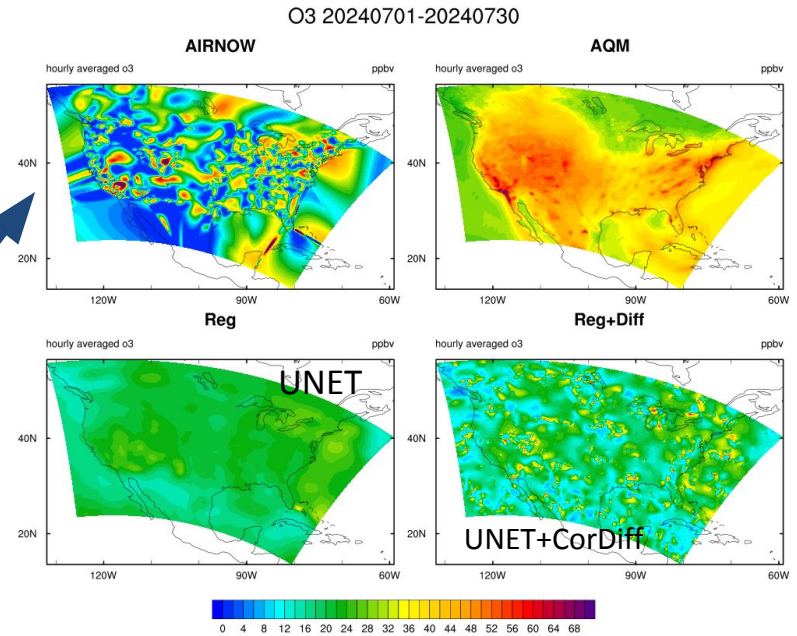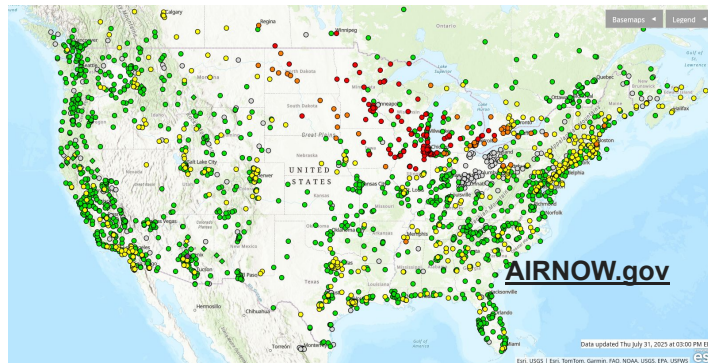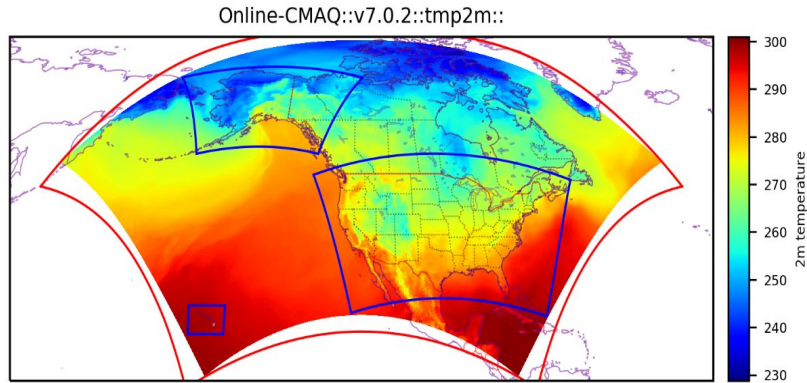- **OCELOT builds the capability with NOAA datasets + modular GNN**

Bias Correction- GFS CAPE

CAPE fhr=24 ave(IC=20230601-20230831)

Generated by using the 400,000th sampling step in regression and 500,000th sampling step in diffusion

Bias Correction and Downscaling Air Quality Model Prediction - O3
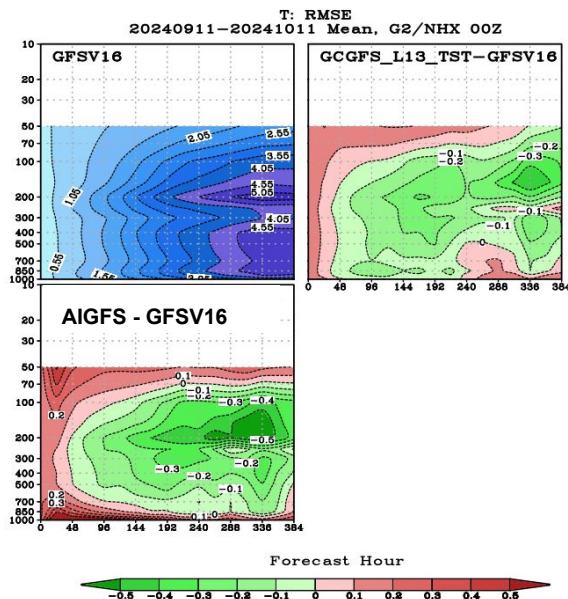## So far not a very successful story

The End

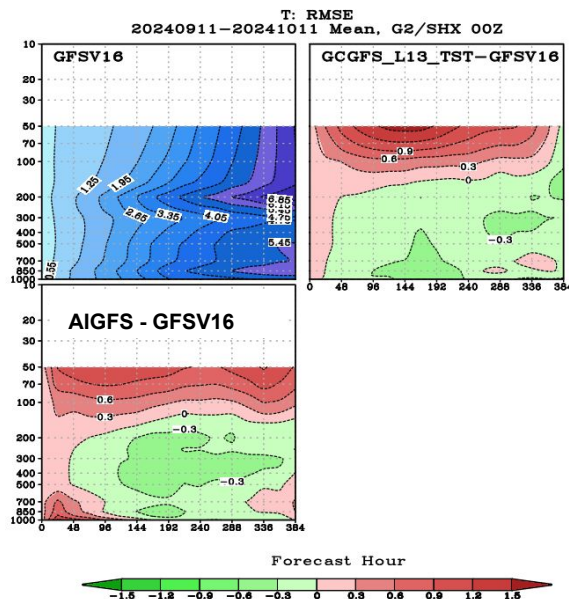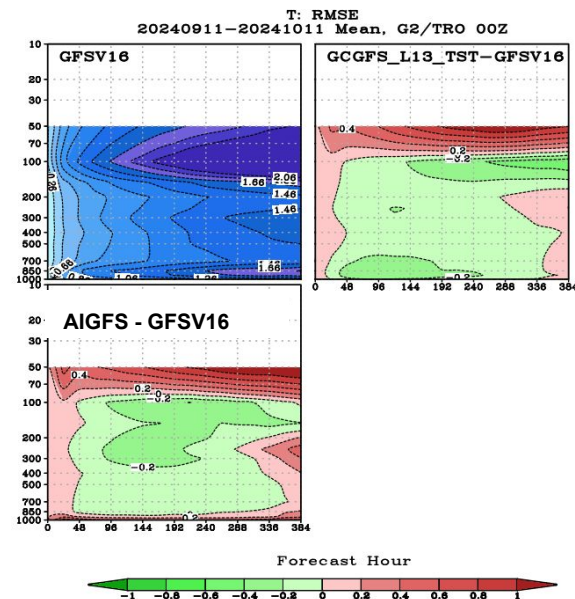# AIGFS Evaluation:

Vertical structure: RMSE TMP

- AIGFS has better performance in most of the troposphere at lead time > 2 days
- AIGFS has a larger RMSE at short lead times and at the top of the atmosphere.